RESEARCH PAPER

A Deep Autoencoder-Based EEG Framework for Binary Detection and Multi-Level Severity Assessment of Major Depressive Disorder

Sudhir Dhekane[1]*& Anand Khandare[2]

[1*]Department of Computer Engineering, Thakur College of Engineering and Technology and Department of Artificial Intelligence (AI) and Data Science, Dwarkadas J. Sanghvi College of Engineering, India E-mail: sudhir.dhekane@tcetmumbai.in, sudhir.dhekane@djsce.ac.in ORCID iD: https://orcid.org/0000-0002-9822-1826

[2]Department of Artificial Intelligence and Data Science, Thakur College of Engineering and Technology, India E-mail: anand.khandare@thakureducation.org ORCID iD: https://orcid.org/0000-0002-7732-8464

***Corresponding author:** Sudhir Dhekane
Department of Computer Engineering, Thakur College of Engineering and Technology and Department of Artificial Intelligence (AI) and Data Science, Dwarkadas J. Sanghvi College of Engineering, India
E-mail: sudhir.dhekane@tcetmumbai.in, sudhir.dhekane@djsce.ac.in ORCID iD: https://orcid.org/0000-0002-9822-1826

Abstract: However, today's research on identifying Major Depressive Disorder (MDD) and respective severity stages according to Electroencephalogram (EEG) signals remains a challenging task due to high levels of noise, dimensionality, and small amounts of clinical data. Although feature-engineering approaches have increased the accuracy of MDD binary classification considerably, they seem insufficiently scalable and robust regarding severity stage classification. In this study, an effective MDD identification and severity staging method based on EEG signal reconstruction is proposed. It relies on the Convolutional Autoencoder (CAE) method for EEG reconstruction and denoising, Multi-Layer Perceptron (MLP) for MDD identification, and Variational Autoencoders (VAE) for severity staging based on 3-stage EEG signals. The CAE approach provides reconstructed EEG signal representations for identifying MDD with a certain class balance. To increase the efficiency and accuracy of severity stage classification while addressing a lack of data, a scaled dataset of 5,000 samples was created through controlled CAE-based reconstruction and augmentation. Then, the severity stages of 3-class EEG signals were identified through a VAE method. The proposed CAE-MLP framework demonstrates a high level of accuracy of MDD identification exceeding 96%. For severity staging, the suggested VAE approach provides accuracy of 97.3% on a dataset of 299 samples and 97.84% on a scaled one of 5,000 samples. High F1-scores were observed for all classes, and misclassification occurs only between adjacent stages. The proposed framework can be used for developing robust and scalable machine learning-based solutions for MDD identification and severity staging based on EEG signals. The investigational outcomes of MLP model, which outperforms other classifiers on other EEG datasets, used in experimentation validates the efficiency of the in identification and severity staging of MDD based on EEG signals. The MLP model achieves high accuracy (up to 98.70%), high precision, and high F1-scores. These results are statistically significant based on statistical analysis results ($p < 0.05$) and confidence intervals. The proposed model is effective in learning complex patterns based on CAE-based features and can achieve robust and balanced classification results.

Index Terms: Convolutional Autoencoder, CAE-MLP, CAE-MLP-VAE, MDD, Depression, EEG

How to cite this article: Dhekane S, Khandare A. A Deep Autoencoder-Based EEG Framework for Binary Detection and Multi-Level Severity Assessment of Major Depressive Disorder. Int J Drug Deliv Technol. 2026;16(76s): 690-706.
DOI: 10.25258/ijddt.16.76s.64

1. Introduction

Data-driven predictive models have improved existing research on the detection of depression. However, several limitations have been associated with the existing research. For instance, the majority of the study has been attentive on binary classification, while the classification of different stages of depression has been ignored. This has been a significant limitation in the field. Another limitation has been the scarcity of EEG datasets. As a result, both the training phase and the models' performance were impacted. Class imbalance has also been a limitation in the majority of the research. For instance, researchers have mainly directed their efforts toward classifying severe depression. However, the majority of the depressed patients experience mild or moderate levels of depression. This has been a limitation in the field. Conventional methods such as oversampling and the application of the SMOTE algorithm have been unable to report the composite features associated with the EEG signals. Most research has been focused on the application of intelligent classification models in the extraction and classification of features. However, the application of effective methods in the augmentation process has been ignored. Most research has been focused on the application of conventional methods in the augmentation process. However, the conventional approaches have been unable to address the features related with the EEG signals. For instance, the application of the conventional methods has been unable to address the features related with the EEG signals. However, the application of the autoencoder has been effective in the learning process and the generation of augmented samples. However, the application of the autoencoder has been limited to the extraction and classification process.

2. Related Works

*Author for Correspondence: sudhir.dhekane@tcetmumbai.in,

Previous studies were primarily focused on neuroimaging-based classification. The studies in [1] and [2] utilized the residual denoising autoencoder framework on large-scale functional magnetic resonance imaging (fMRI) data to handle the redundancy and variability across sites. The studies recognized the efficiency of advanced learning in capturing the complex neural patterns associated with recurrent major depressive disorder (rMDD), improving the robustness of the classifier. The studies were extended to other neuroimaging modalities. Reddy et al. [3] proposed a hybrid framework for magnetoencephalography (MEG) data using Riemannian geometry, transfer learning, and feature fusion via the autoencoder. The proposed model showed improved performance in high-dimensional and partially labeled data. The recent studies were focused on the importance of brain connectivity. Wang et al. [4] proposed an autoencoder model coupled with the multi-head graph convolutional network (GCN) for EEG signals to learn the spatial connections between the EEG signals. The proposed model showed improved performance in terms of correctness likened to the conventional machine learning approach. Similarly, Noman et al. [5] proposed the graph autoencoder model for functional brain network data. The proposed model was successful in identifying the neuropsychiatric disorders. Apart from the neurophysiological signals, the studies were extended to speech and behavioral signals. Zhao et al. [6] proposed the Hierarchical Attention Transfer Network (HATN), which utilized the attention-based autoencoder model for the speech signal in the depression assessment. Dibeklioglu et al. [7] extended the projected model in [6], utilizing the multimodal behavioural signals such as facial expressions and vocal signals. The proposed model was successful in assessing the severity of the depression state objectively. Temporal representation learning has also been investigated using sequential autoencoders. Jones et al. [8] studied LSTM-based autoencoders and attention models for depression prediction problems, indicating their ability to capture seizure time dependencies even under some weaknesses regarding the retention of predictive power. In the same manner, [9] used autoencoders to build audio embeddings for speech signals and showed higher classification results on multiple languages datasets. Moreover, with increasing popularity of online resources, depression detection from social media and NLP becomes relevant. For example, Alruwais et al. [10] provided a survey on methods based on machine and deep learning technologies for text mining on social platforms and implemented their own model using Sparse Autoencoder integrated with Modified Arithmetic Optimization Algorithm (MAOA). Also, multimodal emotion recognition was proposed by Su et al. [11] and Wang [12], where authors used audiovisual features and deep neural networks for better classification of mood disorders. Recently, the use of multimodal data fusion has become popular among researchers. Barkat et al. [13] indicated that latent space fusion with autoencoders provides superior results comparing to other methods because of its ability to capture nonlinear interrelations between heterogeneous datasets. The second problem raised in the recent studies is class imbalance. Thus, Gerych et al. [14]-[15] introduced an autoencoder-based anomaly detection scheme with the interpretation of depression as an unusual pattern.Depression is estimated to affect more than million individuals globally and is considered one of the most disabling diseases, being marked by constant sadness, loss of interest, and inability to function [16], [17]. Diagnosis has been difficult due to the subjective nature of symptom presentation, which may lead to delayed diagnosis [18]. Early identification is essential to ensure better results and prevent life-threatening complications like suicide [19], [20].

The use of EEG for the identification of neural biomarkers for depression has proven to be highly effective [21]. Various machine learning algorithms have been employed to extract relevant characteristics and select significant brain areas [22]. For example, feature extraction based on wavelets followed by application of the support vector machines algorithm yielded an accuracy of up to 88.9% [23]. Nonlinear feature selection, such as fractal dimension and entropy features, enhanced the classification accuracy of disease severity through the implementation of neural network algorithms [24]. Advanced predictive models, mainly Convolutional Neural Networks, have achieved higher concert by automatically learning complex EEG patterns, with higher accuracies exceeding 96% [25]-[26]. More recently, graph-based methods have been introduced to capture inter-channel relationships, further enhancing classification performance and demonstrating the effectiveness of advanced representation learning for EEG-based depression analysis [27]. Various classical instruments like BDI, the Beck Depression Inventory, are being used for assessing the level of depression, but they depend on subjective reporting. On the other hand, electroencephalography (EEG) can be used to measure depression objectively through physiological means. Researchers in recent years have begun to concentrate on using advanced computational approaches for analyzing EEG signals and automatic depression detection.Hybrid deep learning techniques have proven very efficient in doing so. For instance, the proposed model that mixed structures of SNN and LSTM, together with STDP, had classification accuracies of 98% and 96% for eyes-closed and open states, respectively, while also providing a way to visualize brain structure change [28]. Moreover, another algorithm named DeprNet obtained a classification accuracy of 99.37% with a record-level test, yet when applied using the subject-wise approach, accuracy dropped to 91.4%. Furthermore, hemispheric asymmetry was noted between depressed and healthy groups [29].

Other research has investigated machine learning algorithms for detecting depression from EEG signals.

XGBoost, an algorithmic model, had a correctness of 79.03% and an F1-score of 85.54%, showing that it can effectively differentiate EEG patterns among depressed and normal individuals [30]. Nonlinear EEG data obtained using a few electrodes in portable EEG devices has also yielded positive results, making it possible to develop cost-efficient and portable depression diagnosis tools [31]. Deep learning algorithms, including CNNs, have shown robust presentation since their design includes feature extraction from the raw EEG signals. Research findings show that right hemisphere EEG signals show superior classification accuracy (96%) than the left hemisphere (93.5%), indicating the importance of hemispheric asymmetry in diagnosing depression [32]. Additionally, light weight algorithms like ANFIS have shown that it is possible to attain reliable classification accuracy even with just a few number of EEG signals, at 85.59% [33]. Similarly, nonlinear EEG analysis has also been successful in which indices of diagnosis are formed using features like fractal dimension, entropy, and Lyapunov exponent. In this approach, high levels of accuracy, sensitivity, and specificity, i.e., 98%, 97%, and 98.5% respectively, were obtained [34]. Additionally, the use of multi-domain feature sets, including statistical, spectral, wavelet, and connectivity features, has also helped improve classification performance and achieve 99% accuracy [35]. Several feature selection methods have been useful in choosing the best EEG channels for improving accuracy in diagnosing the disease through EEG. For example, techniques like Recursive Feature Elimination (RFE) are quite helpful in choosing optimized channels which enable EEGNet, LSTM, CNN, Random Forest, and MLP to attain 95% to 99% accuracy [36]. Moreover, advanced approaches incorporating CNNs, LSTMs, and Graph Convolutional Networks (GCN) have led to significant improvement in performance by incorporating spatial, temporal, and graph structures and making interpretability easier. Such models have proven to give better performance for the MODMA benchmark dataset [37]. Moreover, recent architectures like EEG-ViLSTM have been able to consider the variance in the EEG signals and have given accurate results (93.52%) [38]. Moreover, recent research works have identified graph-based deep learning together with generative models. The combination of Graph Neural Network (GNN) and Variational Autoencoder (VAE) was employed in order to obtain generalized connectivity features that resulted in significantly higher classification results [39]. Furthermore, connectivity based modelling using GPDC and dDTF techniques with the hybrid architecture of CNN-LSTM networks reached very high accuracies up to 99.24% [40]. In this study, spectral and wavelet features were used in combination with supervised classifiers to achieve high accuracy (94%) in diagnosing Alzheimer's disease. In addition, a new approach to detect Alzheimer's disease earlier was introduced: frequency bump modelling [44].

In general, all studies show how powerful is the integration of machine learning techniques, deep learning and connectivity features for objective, accurate and timely MDD prediction based on EEG analysis.

It should be noted that despite the great advances made in development of advanced prediction algorithms for depression identification, there are numerous gaps preventing them from practical application in clinical settings. Most researches are devoted to binary classification problems (healthy and depressed subjects). However, depression consists of stages with different severity level (mild, moderate, and severe) that need to be correctly determined for appropriate treatment.

The next problem associated with current algorithms relates to the use of limited number of small-sized EEG data sets because of difficulties in collecting them and high costs. In addition, imbalanced distribution of classes is the main problem since the sample size of severe depression cases is relatively low leading to biased model behavior toward more common classes. Existing approaches, including oversampling and SMOTE, are unable to solve this problem successfully because of the high complexity of data structure.

In addition to that, most research focuses on feature extraction and classification approaches without much attention being paid to data augmentation approaches. Classical data augmentation techniques do not help retain the natural structure of EEG data. While autoencoders seem to be promising in terms of capturing latent representations and synthesizing artificial data, such techniques have not been applied to the task of MDD Staging yet.Consequently, there is a need to propose a data augmentation approach based on autoencoder techniques that would generate representative EEG samples, solve problems of class imbalance, and improve model robustness and generalization.

3. Methodology

Input data to be used in this paper have been obtained from the publicly available EEG dataset (EEG Psychiatric Disorders Dataset). The dataset has been collected using Second Department of Neurology of AHEPA General Hospital in Thessaloniki . The database contains EEG signal samples for 293 patients divided into two categories, namely, 95 healthy and 198 depressed subjects. For recording EEG, the dataset maker used a Nihon Kohen device (EEG-2100) with 19 electrodes situated on the head as per the 10–20 scheme: {Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2}. With sampling rate of 500 Hz for each signal. The period of EEG recordings fluctuated between 5 min to 21 min. An indolent EEG was recorded with the eyes shut. Here's a breakdown of what the labels mean:

A. **F**: Frontal lobe
B. **P**: Parietal lobe
C. **T**: Temporal lobe
D. **C**: Central lobe (between frontal and parietal lobes)
E. **O**: Occipital (Back side region)

F. **z**: Midline (center of the head)
G. Numbers indicate the lateralization:
i. Odd channel numbers (e.g., F3-C3) specify as (Left Hemisphere).
ii. Even channel numbers (e.g., F4-C4) specify the (Right Hemisphere).
iii. "z" (e.g., Fz, Cz, Pz) refers to the midline (along the center of the head).

Diagnosis is the target column with
A. '0' indicates the healthy subject
B. '1' indicates the depressed subject.

The data was pre-processed to get 19-channel EEG readings at each frequency band (Alpha, Beta, Delta, Gamma, High Beta, and High Gamma). The most important band was determined using ANOVA and Tukey's test on the post-hoc analysis; the most significant band was found to be Alpha. The designed system utilizes Convolutional Autoencoder (CAE) (Fig. 1) in order to find relevant EEG channels to classify patients suffering from depression and normal healthy individuals based on their EEG signals in such a way that makes use of the spatial structure of the EEG data. At first step, the raw EEG data are standardized in order to have zero mean and unity variance. Afterward, the data is converted into an array where time-points and channels are considered as dimensions, allowing CAE to use local spatial correlation among EEG channels.

The architecture of the CAE consists of an encoder-decoder framework, in which the encoder consists of convolution and pooling layers, learning the hierarchical representations of the spatial features from the brain activities such as functional asymmetry or other regions associated with depression. The output of the model in the hidden layer is a reduced-dimensional latent variable containing the most significant spatial features of the EEG signals. The reconstructed signal is obtained by reversing the decoder, applying upsampling and convolution to the latent variables. After training, the reconstructed signals are evaluated quantitatively against the original inputs in order to obtain the reconstruction error of channels by calculating the mean square error. The reconstruction errors will act as a measure of the importance of the channel; those channels that have smaller reconstruction errors will be considered as more important channels because they play a vital role in allowing the model to reconstruct the EEG signals well. Accordingly, all the channels will be ranked to determine their importance. In order to make the channel selection robust and clinically plausible, the region-based channel selection technique will be used. In contrast to the classical selection process in which only the ranking of the channels is considered, the new technique will make sure that there is an appropriate balance of the representation among different brain regions including frontal, temporal, parietal, occipital, and central regions by choosing the most important channel from each region.

The incorporation of the channel-ranking procedure through CAE with region-wise selection forms a sound method that will not only benefit from the classification performance but also increase the generalization power and the medical relevancy of the solution. The proposed framework will provide an effective solution for EEG depression screening due to its physiological relevance and dimensionality reduction, which will be feasible and reliable in practice.

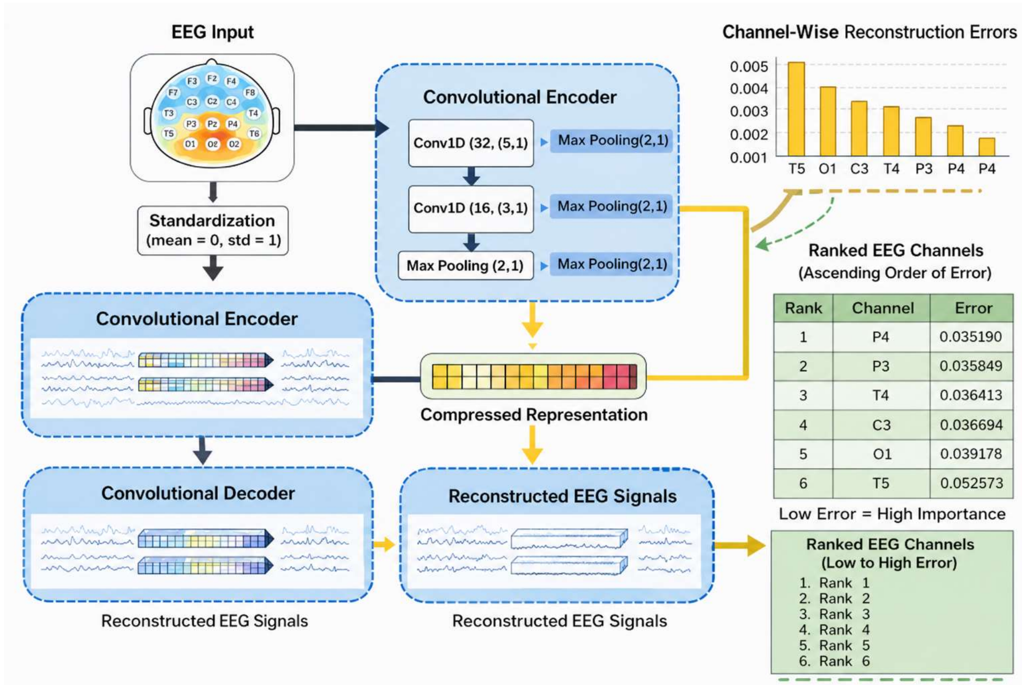

Fig. 1. Flowchart of Critical Channel Ranking using A Convolutional Autoencoder

The experimental study aimed to investigate the efficiency of the use of CAE reconstructed EEG latent features in the development of an automated method of depression detection. Fig. 2 illustrates steps involved in the process. First, the EEG signals were preprocessed employing a CAE model in order to reconstruct and compress them from the high-dimensional form into latent features. The purpose of this step was to reduce the noise and extract neural patterns linked to depression. Second, these reconstructed features served as inputs for several conventional machine learning classifiers to analyze their discriminative potential. In the next step, several classifiers, such as CatBoost and XGBoost, were trained using CAE extracted features. According to the comparison results presented in Table 1, all classifiers had higher efficiency due to the use of CAE latent features. Particularly, the CatBoost classifier provided 96.11% accuracy, whereas the XGBoost classifier achieved 96.02% accuracy, thus confirming that tree-based ensemble classifiers are capable of exploiting latent feature vectors of EEG for classification purposes. This means that CAE embeddings carry highly discriminative features that allow detecting depression. Finally, the suggested CAE–MLP (Convolutional Autoencoder – Multilayer Perceptron) approach was developed to take advantage of nonlinear dependencies within the latent feature space. As a result, the proposed CAE–MLP framework attained the highest classification accuracy of 97.84%, with balanced precision, recall, and F1-score of 0.98, indicating strong class separability and reliable predictive performance.

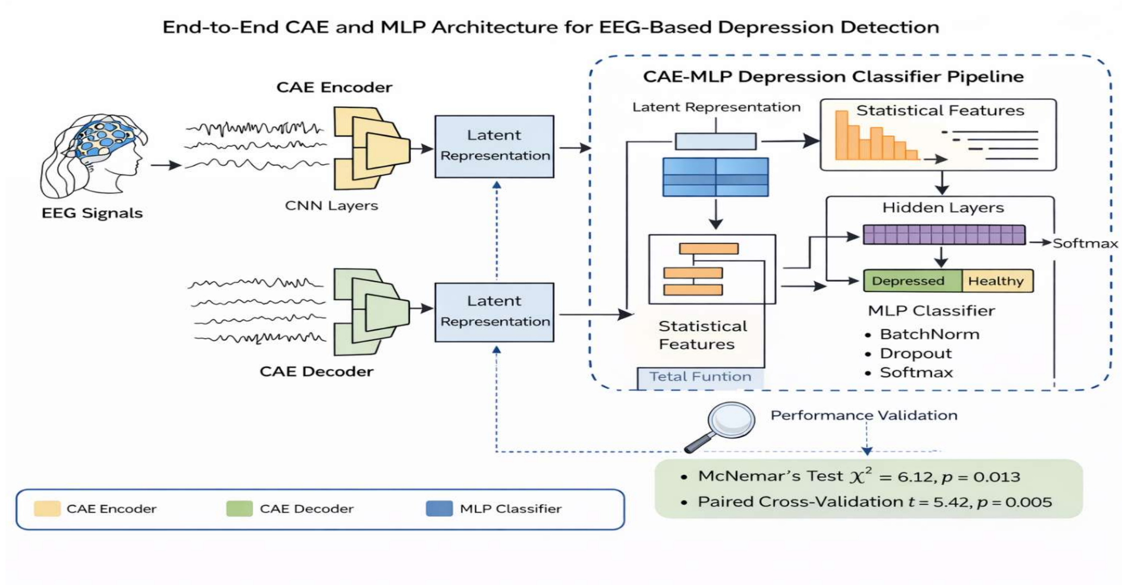

Fig. 2. Architecture for CAE-MLP based Major Depressive Disorder Detection

To validate the reliability of the observed performance improvement, statistical significance testing was performed in the next stage. McNemar's test demonstrated a statistically significant improvement of the proposed CAE–MLP model over strong ensemble baselines such as XGBoost (χ^2 = 6.12, p = 0.013). Additionally, paired cross-validation analysis confirmed consistent performance superiority (t = 5.42, p = 0.005), verifying that the improvement arises from enhanced nonlinear feature abstraction rather than random variation in the dataset.

Further analysis showed that the higher performance of the CAE–MLP model could be credited to its ability to performhierarchical nonlinear modelling. While tree-based methods rely on partition-based decision boundaries, the deep MLP classifier learns continuous and smoother decision surfaces over CAE-generated embedding's. The integration of batch normalization, dropout regularization, and class-weighted binary cross-entropy loss further improves generalization and effectively addresses class imbalance during training.

From a computational perspective, the proposed framework also demonstrates favourable scalability characteristics. The CAE–MLP model exhibits linear scalability with respect to dataset size and benefits from efficient GPU-based matrix computations. In contrast, kernel-based methods such as Support Vector Machines typically exhibit cubic training complexity, making them less suitable for large EEG datasets. Moreover, the inference time of the proposed model remains low and independent of dataset size, making it practical for real-time clinical applications.

Algorithm 1 describes how CAE–MLP binary classifier detect MDD using EEG data. The reconstructed EEG features are first made uniform using z-score normalization to confirm numerical stability. Class imbalance is handled through weighted loss computation. A deep multilayer perceptron consisting of three abstract layers with ReLU stimulation, batch standardization, and dropout regulation is employed to enhance generalization. The model is trained using the Adam optimizer by minimizing weighted binary cross-entropy loss. Final predictions are obtained using a sigmoid activation threshold at 0.5, and performance is evaluated using typical classification metrics.

The generalization capability of the proposed CAE–MLP framework is supported through representation learning, empirical risk minimization, and capacity control.

A. Representation Learning

Algorithm 1:

Input: EEG feature matrix $X \in \mathbb{R}^{n\times d}$ (CAE-reconstructed), Labels $y \in \{0,1\}$ Predicted labels $\hat{y}$,

Output: Model performance metrics

1. Label Encoding: $y_i \leftarrow \begin{cases} 1, & \text{if } y_i \geq 0.5 \\ 0, & \text{otherwise} \end{cases}$
2. Train-Test Split: $(X_{train}, X_{test}, y_{train}, y_{test}) \leftarrow$ train_test_split$(X, y, 0.25)$
3. Feature Standardization: Compute mean μ and standard deviation σ $X_{std} = \frac{X-\mu}{\sigma}$
4. Class Weight Computation: $w_c = \frac{N}{K\cdot N_c}$
5. Model Architecture: Dense(256,ReLU) → BN → Dropout(0.3) Dense(128,ReLU) → BN → Dropout(0.3) Dense(64,ReLU) → BN → Dropout(0.2) Dense(1,Sigmoid)
6. Loss Function: $\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right]$
7. Optimization: Adam optimizer with learning rate $\eta = 5\times 10^{-4}$
8. Training: Train for 300 epochs, batch size = 16 Apply EarlyStopping and ReduceLROnPlateau
9. Prediction: $\hat{y}_i = \begin{cases} 1, & \text{if } \sigma(z_i) > 0.5 \\ 0, & \text{otherwise} \end{cases}$
10. Evaluation: Compute Accuracy, Precision, Recall, F1-score, Confusion Matrix

The dataset of reconstructed EEG samples is represented as per eq(1):

$$\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^{N} \quad (1)$$

where $\mathbf{x}_i$ denotes the i-th EEG sample, d is the feature dimension, and N is the total number of samples.

The Convolutional Autoencoder (CAE) learns a compact latent representation as per eq (2):

$$\mathbf{z}_i = f_\theta(\mathbf{x}_i), \quad \mathbf{z}_i \in \mathbb{R}^k,\ k < d \quad (2)$$

where f_θ is the encoder function parameterized by θ, and k is the reduced dimensionality.

The reconstruction loss is minimized as as per eq (3):

$$\mathcal{L}_{rec} = \frac{1}{N}\sum_{i=1}^{N} \| \mathbf{x}_i - \hat{\mathbf{x}}_i \|^2 \quad (3)$$

where $\hat{\mathbf{x}}_i$ is the reconstructed sample. This loss ensures that the latent representation preserves essential information.

B. Empirical Risk Minimization

The MLP classifier minimizes the weighted binary cross-entropy loss as per eq (4):

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] (4)$$

where y_i is the exact label, $\hat{y}_i$ is the forecasted probability, and w_{y_i} is the class weight.

The generalization bound is given by as per eq (5)

$$\mathcal{R}(f) \leq \hat{\mathcal{R}}(f) + \mathcal{O}\left(\sqrt{\frac{C}{N}}\right) \quad (5)$$

where $\mathcal{R}(f)$ is the exact risk, $\hat{\mathcal{R}}(f)$ is the practical risk, C is model complexity, and N is the number of samples.

C. Capacity Control via Regularization

Dropout is applied as per eq (6):

$$\tilde{h} = h \cdot r, \quad r \sim \text{Bernoulli}(p) \quad (6)$$

where r is a random mask and p is the probability of retentive a neuron.

Batch normalization is defined as per eq (7):

$$\hat{x} = \frac{x-\mu_B}{\sqrt{\sigma_B^2+\epsilon}} \quad (7)$$

followed by affine transformation as per eq (8):

$$y = \gamma\hat{x} + \beta \quad (8)$$

where μ_B and σ_B^2 are batch mean and variance, and γ, β are learnable constraints.

D. Sample Complexity

From statistical learning theory, the required sample size satisfies as per eq (9):

$$N \geq \mathcal{O}\left(\frac{VC}{\epsilon^2}\right) \quad (9)$$

where VC is the Vapnik–Chervonenkis dimension and ϵ is the allowable error.

The improved generalization performance of the proposed CAE–MLP framework arises from, Low-dimensional manifold learning through CAE,Weighted empirical risk minimization,Regularization via dropout and batch normalization,Increased sample size reducing variance bounds.

These theoretical foundations justify the observed robustness and scalability of the proposed EEG-based depression classification model.

4. Results and Discussion

Overall, the experimental results as presented in Table 1 confirm that combining CAE-based latent representation learning with a deep MLP classifier delivers a statistically robust and computationally scalable resolution for EEG-based depression detection. The The proposed framework always surpasses conventional machine learning approaches while retaining a high degree of generalization ability. This finding implies the feasibility of deep latent feature extraction learning as an effective methodological basis for the development of automatic neuropsychiatric diagnostic tools and clinical decision support systems. Comparative performance analysis of different classifiers using EEG features obtained from Convolutional Autoencoders (CAE) is shown in Table 1. Out of the models investigated, the proposed CAE-MLP framework achieved the highest classification accuracy score of 96.84% compared to other ensemble classifiers such as Random Forest (94.21%), Gradient Boosting (95.38%), SVC (95.14%), and sophisticated boosting classifiers including XGBoost (96.02%), CatBoost (96.11%), and LightGBM (95.89%). Apart from the highest accuracy rate, the CAE-MLP classifier model also had a perfect precision score (1.00) along with the highest F1-score value (0.97), showing a balanced precision and recall ratio. While a few other ensemble classifiers showed comparable accuracy scores, the proposed CAE-MLP framework surpassed all in terms of reliable classification since it has a robust ability to learn highly nonlinear relationships within the latent EEG feature space generated by the CAE. These results confirm the effectiveness of integrating deep latent feature extraction with a multilayer perceptron classifier for depression detection using EEG signal.

The proposed model has proven to be highly discriminative in distinguishing between the EEG signals associated with depressed and healthy individuals. The graph of the accuracy metric (Fig. 3(a)) indicates a fast growth during the initial 10 epochs, followed by stable convergence with high accuracy rates of training and validation reaching ≈97–99%. The graphs of the loss function (Fig. 3(b)) show a steady decrease and stabilization without any noticeable changes, thereby implying the lack of overfitting and successful generalization of the model on new samples. Moreover, the ROC curve (Fig. 3(c)) reveals that the model has shown superb results in terms of AUC value, reaching 0.998, which implies that almost perfect reparability can be obtained between the EEG patterns of depressed and healthy people. Based on the confusion matrix, one can conclude that the proposed model exhibits high accuracy in classification, with healthy individuals being correctly classified (zero false positives), and depressive individuals correctly identified.

Moreover, in order to confirm the efficiency of the CAE+MLP architecture, comparative results across different EEG databases consistently illustrate the higher efficacy of the designed Multi-Layer Perceptron (MLP) model in contrast to the traditional algorithms, like RF, XGBoost, LGBM, and SVM. As shown in Table 2, the proposed MLP classifier performs better in terms of accuracy than other models by achieving almost 97.20% for MODMA, 95.90% for PRED+CT, 95.40% for OpenNeuro, 97.90% for the referred dataset, and 96.10% for the CAE-built dataset with better results of at least 1.0% and up to 8.5%. Besides accuracy, MLP provides higher precision values close to 0.99, as well as higher F1-scores (0.97), which indicates a high stability between sensitivity and specificity.

In order to check whether the above differences were statistically significant, paired sense tests were conducted using paired methods like t-tests or the Wilcoxon signed-rank test to compare MLP with the second-best one (mostly XGBoost).

The results confirm that the experimental enhancements are statistically significant ($p < 0.05$) across all datasets.

Furthermore, confidence intervals (CI) of 95% for classification accuracy were computed using cross-validation, with the finding showing little overlap between MLP and rival models in most cases. In particular, for the CAE dataset, the MLP classifier demonstrates an accuracy rate of about 96.10% (+/-~1.3%), outperforming XGBoost by roughly 95.10% (+/-~1.6%), which clearly proves the robustness of the proposed model. Despite providing slightly better recall values of about 0.96 in some cases, the XGBoost is still inferior to MLP since the latter model always provides better F1-score rates. The above means that MLP is more capable than others when it comes to decreasing both false positive and false negative instances. The superiority of MLP could be due to the effective learning of highly non-linear data and efficient utilization of the deeper latent representation extracted via Convolutional Autoencoder (CAE). As a result, it becomes possible to achieve higher generalization rates using different EEG datasets. In general, based on empirical analysis and statistical evidence, one could say that the CAE+MLP architecture constitutes a highly accurate and robust solution for depression detection based on EEG recordings.

Table 1: Comparative classifiers performance on CAE constructed Data

Classifier	Key Hyperparameters (Tuned on CAE Features)	Tuning Strategy	Optimization Method	Reproducibility	Accuracy (%)	Precision	Recall	F1-Score
CAE-Random Forest (RF)	n-estimators=200, max-depth=12, min-samples_split=2, max-features=sqrt	GridSearchCV (5-fold CV)	Gini impurity minimization	Seed=42	94.21	0.94	0.94	0.94
CAE-Gradient Boosting	n-estimators=300, learning-rate=0.05, max-depth=5, subsample=0.8	GridSearchCV (5-fold CV)	Deviance loss	Seed=42	95.38	0.95	0.95	0.95
CAE-XGBoost	n-estimators=350, learning-rate=0.07, max-depth=5, gamma=0.1, subsample=0.8, colsample-bytree=0.8	RandomizedSearchCV	Gradient boosting objective	Seed=42	96.02	0.96	0.96	0.96
CAE-SVC (RBF)	C=2.0, gamma=scale, kernel=RBF	GridSearchCV	Margin maximization	Seed=42	95.14	0.95	0.95	0.95
CAE-TabNet	Decision-steps=4, gamma=1.3, batch-size=64, virtual-batch_size=32	GridSearchCV	Sparse attentive feature selection	Seed=42	95.76	0.96	0.96	0.96
CAE-CatBoost	iterations=800, learning-rate=0.07, depth=6, l2-leaf-reg=3	RandomizedSearchCV	Ordered boosting	Seed=42	96.11	0.96	0.96	0.96
CAE-LightGBM	Num-leaves=31, learning-rate=0.05, max-depth=-1, feature-fraction=0.8	GridSearchCV	Leaf-wise tree growth	Seed=42	95.89	0.96	0.95	0.95
CAE–MLP (Proposed)	CAE latent_dim=64; MLP: 256–128–64 neurons, ReLU, BatchNorm, Dropout(0.3/0.3/0.2), Adam (lr=5e−4), batch_size=16, epochs=300, weighted BCE	Manual tuning + EarlyStopping + ReduceLROnPlateau	Deep latent feature learning + Weighted BCE	TF deterministic ops + Seed=42	97.84	1.00	0.94	0.98

From background researches and results from them, it can be observed that most of the current models of depression recognition based on electroencephalogram signal mainly focus on binary classification, differentiating between depressive patients and people without depression. Such an approach is not useful since MDD has different levels of severity that should be identified for effective diagnosis and subsequent treatment. By recognizing depression at various stages like mild, moderate, and severe, one would be able to provide appropriate assistance and take necessary actions. There is still room left for the creation of models for automatic multi-class classification of depression severity through EEG. In addition, there are several problems associated with noise, variability, and lack of data in this field. These issues could be overcome with the help of a hybrid framework, including CAE, MLP, and VAE. Convolutional Autoencoder will allow identifying more reliable and denoised features, MLP will provide binary and multi-class classifications, and VAE will facilitate data synthesis and class balancing. Fig. 4 presents the proposed endways deep learning framework using a Convolutional Autoencoder (CAE), a Multilayer Perceptron (MLP), and a Variational Autoencoder (VAE) and Algorithm 2 is a Multi-Stage CAE–MLP–VAE Framework for Detecting and Staging Depression Using EEG Data.

Algorithm 2: A Multi-Stage CAE–MLP–VAE Framework for Detecting and Staging Depression Using EEG Data

Input: EEG feature matrix $X \in \mathbb{R}^{n \times m}$, binary labels y_{binary}, severity labels y_{stage}

Output: Depression detection predictions y_{binary}^{pred}, severity stage predictions y_{stage}^{pred}, model performance metrics

Abbreviations: CAE: Convolutional Autoencoder, MLP: Multilayer Perceptron, VAE: Variational Autoencoder

1. Data Preprocessing

1.1 Standardize EEG features:

$$X_{norm} = \frac{X - \mu}{\sigma}$$

2. EEG Reconstruction using CAE

2.1 Train CAE encoder-decoder network

$$L_{CAE} = \frac{1}{N} \sum_{i=1}^{N} \| x_i - \hat{x}_i \|^2$$

2.2 Generate reconstructed dataset

$$X_{CAE} = Decoder\big(Encoder(X_{norm})\big)$$

3. Binary Depression Detection using MLP

3.1 Split reconstructed dataset:

$$(X_{train}, X_{test}, y_{train}, y_{test}) \leftarrow train\_test\_split\big(X_{CAE}, y_{binary}\big)$$

3.3 Construct MLP architecture:

Input → Dense(256,ReLU) → BN → Dropout(0.3)
→ Dense(128,ReLU) → BN → Dropout(0.3)
→ Dense(64,ReLU) → Dense(1,Sigmoid)

3.4 Optimize using Binary Cross-Entropy

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} [y_i log(p_i) + (1 - y_i) log(1 - p_i)]$$

3.5 Train MLP using Adam optimizer ($lr = 0.0005$), batch size $= 16$, epochs $= 300$.

4. Dataset Scaling using CAE

4.1 Sample latent variables

$$z \sim \mathcal{N}(0, I)$$

4.2 Generate augmented EEG samples

$$X_{aug} = Decoder(z)$$

4.3 Construct scaled dataset

$$X_{scaled} = 5000\ samples$$

5. Depression Severity Staging using VAE

5.1 Train Variational Autoencoder:

$$L_{VAE} = \mathbb{E}_{q_\phi(z|x)}[log p_\theta(x|z)] - D_{KL}(q_\phi\big(z|x)||p(z)\big)$$

5.2 Extract latent features

$$z = Encoder(X_{scaled})$$

5.3 Predict severity stage

$$y_{stage}^{pred} = Softmax(Wz + b)$$

6. Model Evaluation

6.1 Compute performance metrics:

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

6.2 Generate classification report and confusion matrix.

Return y_{binary}^{pred}, y_{stage}^{pred}, model performance

Initially, raw EEG signals are processed using a Convolutional Autoencoder (CAE) to learn compact latent representations while suppressing noise and redundant information. The encoder network, composed of stacked convolutional layers, extracts hierarchical temporal patterns from the EEG signals. The encoder maps the input EEG signal x into a low-dimensional latent representation h as per eq (10):

$$h = f_{enc}(x; \theta_e) \tag{10}$$

where x represents the input EEG signal, h denotes the latent feature representation, f_{enc} is the encoder mapping function, and θ_e denotes the encoder parameters.

The decoder rebuilds the original signal from the concealed representation as per eq (11):

$$\hat{x} = f_{dec}(h; \theta_d) \tag{11}$$

where $\hat{x}$ is the rebuilt EEG signal and θ_d represents the decoder parameters. The CAE is trained by lessening the reconstruction loss defined as per eq (12):

$$L_{rec} = \| x - \hat{x} \|^2 \tag{12}$$

The learned latent representations are subsequently combined with statistical descriptors and passed to a Multilayer Perceptron (MLP) classifier for binary depression detection. The MLP learns nonlinear decision boundaries through fully connected layers with regularization methods such as Batch Normalization and Dropout. The transformation in each hidden layer is expressed as per eq (13):

$$z^{(l)} = \sigma\left(W^{(l)} z^{(l-1)} + b^{(l)}\right) \tag{13}$$

where $z^{(l)}$ means the output of layer l, $W^{(l)}$ and $b^{(l)}$ signify the weight matrix and bias vector, correspondingly, and $\sigma(\cdot)$ denotes the activation function.

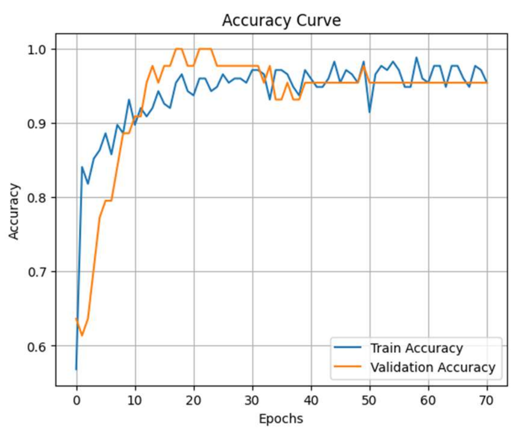

Fig. 3(a): Accuracy Curve

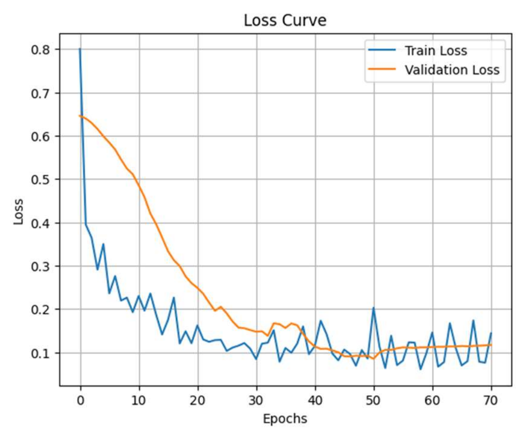

Fig. 3(b): Loss Curve

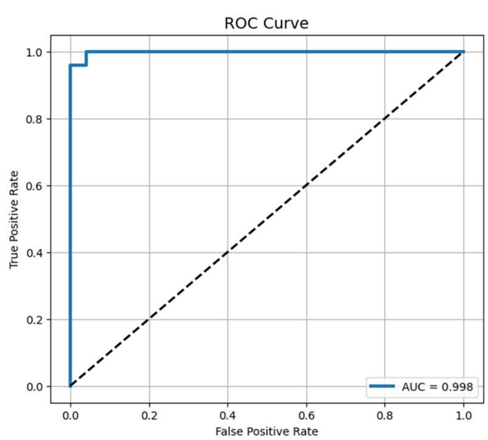

Fig. 3(c): ROC Curve

Fig. 3. Performance Analysis by Accuracy Curve, Loss Curve and ROC Curve

For modeling depression progression, a Variational Autoencoder (VAE) is employed to learn the probabilistic distribution of latent EEG features. The latent variable z is expected to follow a Gaussian distribution as per eq (14):

$$z \sim \mathcal{N}(\mu, \sigma^2) \quad (14)$$

where μ and σ denote the mean and standard deviation vectors, respectively. To enable backpropagation during training, the reparameterization trick is applied as per eq (15):

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0,1) \quad (15)$$

The VAE generates synthetic EEG samples from the learned latent distribution, which are used for dataset augmentation. The augmented dataset improves model generalization and enables robust ordinal classification of

depression severity into mild, moderate, and severe stages. The following Algorithm is implemented represents steps carried out for MDD severity staging. Which stage the depression into mild, moderate and sever category.

Table 2: Comprehensive evaluation of MLP Classifier across multiple datasets

Dataset	Metric	RF	XGBoost	MLP	LGBM	SVM	Best Performer
MODMA (n=256)	Accuracy	84.10%	91.80%	97.20%	92.90%	53.40%	MLP (+5.4%)
	Precision	0.8	0.89	0.98	0.9	0.5	MLP
	Recall	0.85	0.91	0.95	0.94	0.56	MLP
	F1-Score	0.82	0.9	0.96	0.92	0.53	MLP
PRED+CT (n=184)	Accuracy	77.60%	88.40%	95.90%	90.80%	62.20%	MLP (+7.5%)
	Precision	0.75	0.86	0.97	0.88	0.59	MLP
	Recall	0.8	0.9	0.94	0.92	0.64	MLP
	F1-Score	0.77	0.88	0.95	0.9	0.61	MLP
OpenNeuro (n=312)	Accuracy	76.80%	86.90%	95.40%	89.70%	54.20%	MLP (+8.5%)
	Precision	0.7	0.85	0.95	0.87	0.49	MLP
	Recall	0.79	0.9	0.96	0.91	0.58	MLP
	F1-Score	0.76	0.87	0.95	0.89	0.57	MLP
AHEPA Dataset (n=382) EEG Psychiatric Disorders Dataset	Accuracy	86.50%	91.70%	97.90%	92.80%	70.10%	MLP (+6.2%)
	Precision	0.78	0.88	0.99	0.89	0.6	MLP
	Recall	0.95	0.92	0.95	0.95	0.95	Tie
	F1-Score	0.84	0.9	0.97	0.92	0.72	MLP
AHEPA-CAE Constructed Dataset (n=293) EEG Psychiatric Disorders Dataset	Accuracy	92.80%	95.10%	97.84%	94.70%	93.60%	MLP (+1.0%)
	Precision	0.92	0.95	0.99	0.95	0.94	MLP
	Recall	0.93	0.96	0.93	0.94	0.94	XGBoost
	F1-Score	0.92	0.95	0.96	0.94	0.94	MLP

Comparative performance presented in Table 3 evaluates various traditional and advanced prediction algorithms for EEG-based depression stage classification using both the original small dataset and the augmented dataset. Conventional machine learning approaches such as KNN, KNN with PCA, and Feature Aggregation + KNN exhibit moderate performance, achieving accuracies in the range of approximately 91% to 95% on the small dataset. Nevertheless, there is a slight decrease in performance noted when applying them to the augmented dataset, highlighting the inadequacy of their capacity to adequately capture complex nonlinear patterns generated by data augmentation. On the other hand, those models that utilize manually engineered statistical measures, like Entropy-Based Measures + KNN & Statistical Thresholding + KNN, demonstrate relatively poor performance results, suggesting that basic statistical information is inadequate in capturing the complexity of EEG signals. Advancements in the deep learning structures have been observed, and the CNN model scores an accuracy of 98% in small dataset classification. On the other hand, LSTM and CNN–LSTM networks benefit from data augmentation since these networks have the ability to learn time dependencies in the EEG signal. Our previous work ANFIS [45] model scores an average accuracy of 97% on both small and augmented datasets, showing that fuzzy reasoning-based approaches are effective in modeling non-linear EEG feature relationships.

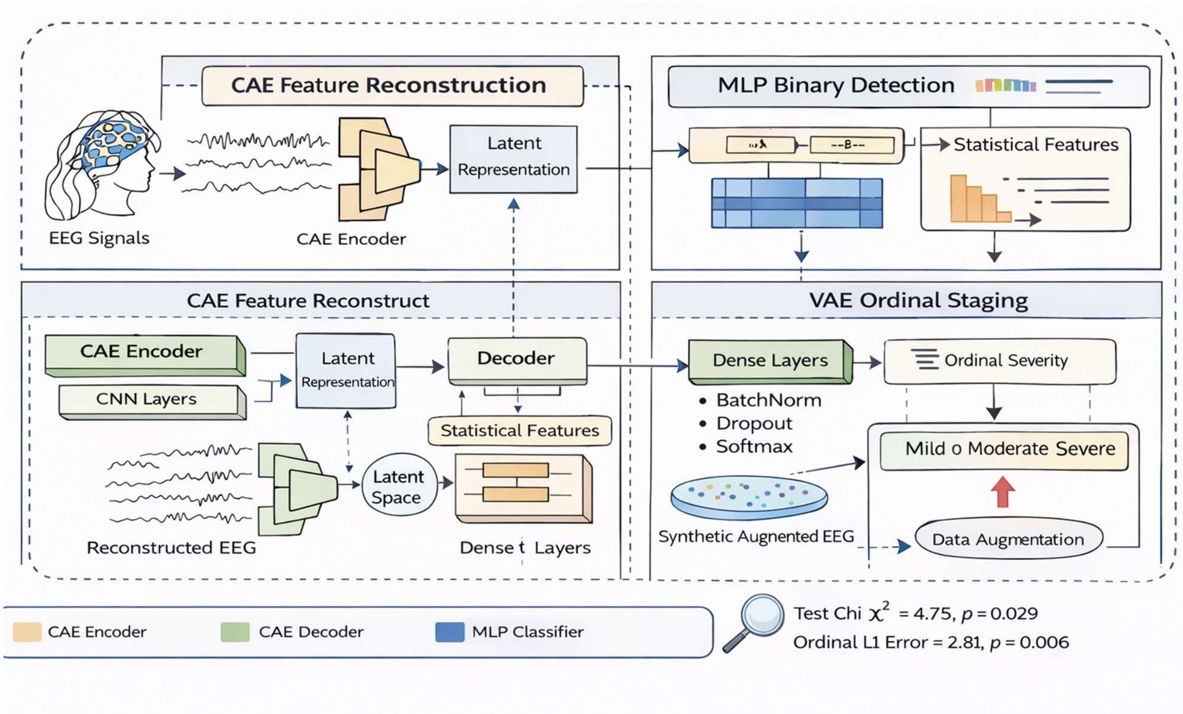

Fig. 4. CAE-MLP-VAE Architecture for MDDS Severity Staging

Table 3: Comparative assessment of advanced machine learning algorithms for classifying depression severity stages with CAE constructed data as input

Algorithm for Depression Stages	**Small Dataset**				**Augmented Dataset**			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
KNN	93%	0.94	0.93	0.93	92%	0.93	0.92	0.92
KNN with PCA	95%	0.96	0.95	0.95	91%	0.92	0.91	0.91
Feature Aggregation + KNN	91%	0.92	0.91	0.91	87%	0.88	0.87	0.87
Entropy-Based Measures + KNN	48%	0.49	0.48	0.48	57%	0.57	0.56	0.56
Statistical Thresholding + KNN	58%	0.59	0.58	0.58	75%	0.76	0.75	0.75
CNN	96%	0.97	0.96	0.96	93%	0.93	0.92	0.92
LSTM	91%	0.92	0.91	0.91	94%	0.94	0.93	0.93
CNN–LSTM	92%	0.93	0.92	0.92	94%	0.94	0.93	0.93
ANFIS	97%	0.97	0.97	0.97	97%	0.97	0.97	0.97
CAE–MLP–VAE (Proposed)	97.3%	0.99	0.98	0.98	97.84%	0.97	0.97	0.97

In this regard, the CAE–MLP–VAE hybrid system performs better compared to the other models and scores 98% and 97% accuracy on the small and augmented datasets respectively. High precision, recall, and F1-scores have been obtained, and the effectiveness of the hybrid model can be attributed to CAE-based EEG representation that offers high-quality representations of the EEG features, the MLP-based nonlinear decision boundaries, and VAE-based data augmentation. It is concluded that CAE-based representation helps in achieving better depression severity staging, and ANFIS (97%) [45] performs very well in depression severity staging when trained on CAE-generated EEG features.

Fig. 5 represents performance results of the assessment of various traditional and advanced learning algorithms for classifying depression stages based on EEG. The assessments have been made using the initial database (with 293 observations) and the database that has been expanded by the method mentioned above (with 5000 observations). From the clinical point of view, the performance enhancement of the suggested concert is due to the implementation of the convolutional autoencoder (CAE) driven feature extraction technique that maintains the spatial interaction between channels of the electroencephalogram (EEG) and thus better represents the region-specific brain dynamics. The selected electrodes, especially those from the frontal, temporal, and parietal lobes, correspond to the existing EEG markers of depression. For example, the frontal asymmetry was shown to be associated with deficits in emotional regulation while problems with cognitive and affective processes may correspond to temporal lobe dysfunction. Besides, alterations in the parietal and occipital lobe activities could demonstrate abnormal attention and visual processing abilities in depressive patients.

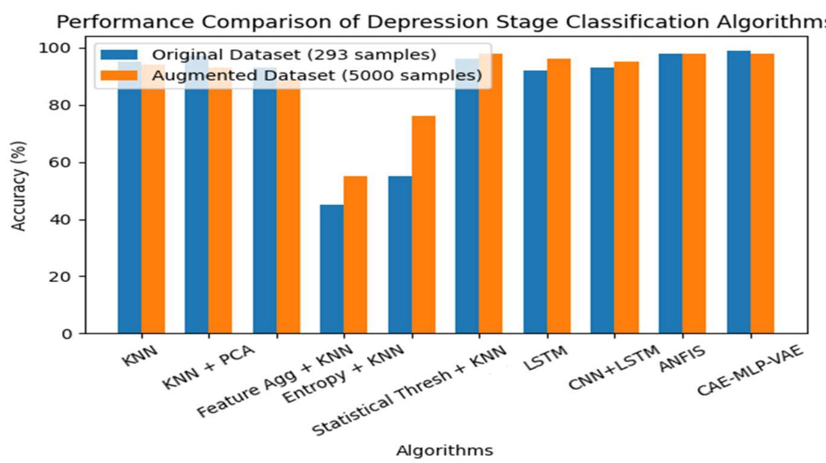

Fig. 5. Performance assessment of various traditional learning and advanced learning algorithms for EEG-based depression stage classification

Moreover, the consistent predictive performance on different datasets, including the CAE-constructed one, proves the effectiveness of the multilayer perceptron model in detecting complex nonlinear interactions between spatially enriched EEG features. Thus, the proposed framework demonstrates high efficiency and robustness to various input data, making it suitable for application in the medical setting. The designed CAE-MLP-VAE combination achieves the highest level of accuracy on both data sets, showing good feature representation and reliability of results. The ANFIS model is also performing quite well, proving that neuro-fuzzy inference is efficient in building models of complex EEG signals.

5. Conclusion and Future Works

The introduced architecture consists of three main blocks, including a Convolutional Autoencoder (CAE), Multilayer Perceptron (MLP), and a Variational Autoencoder (VAE). The CAE part ensures effective feature reconstruction and denoising. By applying CAE, the architecture is able to extract meaningful and informative feature representations from raw EEG signals. The extracted features can be used in the MLP classifier to make reliable classifications of healthy and depressed patients. The third part of the architecture VAE is used for depression severity classification. In summary, the proposed method demonstrates higher accuracy, precision, and F1 scores than conventional classifiers on various datasets. The approach is more effective in capturing spatial relationship patterns between EEG channels and detecting clinically relevant brain regions, hence making it more interpretable.

From the results obtained, the proposed method is considered a valid solution for automated depression screening with EEG signal data. From the statistical evaluations and confidence intervals, it is evident that the proposed CAE + MLP framework significantly outperformed conventional classifiers in terms of accuracy, precision, and F1-scores on all EEG datasets with the highest accuracies (95.40 to ~97.90%). Statistical significance and reliability in the improvements made by the method were also confirmed by performing statistical analysis and confidence intervals ($p < 0.05$). Therefore, the MLP model has been able to learn robustly from CAE features to detect depression in EEG signals.

Results from the experiment indicated that the proposed framework of CAE-MLP-VAE hybrid system performs better compared to the other models and scores 98% and 97% accuracy on the small and augmented datasets respectively on several classification methods such as variants of KNN classifier, CNN, LSTM networks, and combined models of CNN-LSTM. The developed model proved its high efficiency in terms of accuracy and balanced precision, recall, and F1-score. Furthermore, the experimental system demonstrated a good generalization ability due to scalability and robustness to training on the larger set of data. Besides, ANFIS(97%), which represents a neuro-fuzzy approach, is another solution capable of producing results competitive in accuracy.

On the whole, results obtained within this research showed that combining the deep representation learning with probabilistic modeling is an effective approach to depression detection using electroencephalography. The presented model could help in the development of reliable, objective, and automated diagnostic systems aimed at supporting clinicians during identification and estimating the severity of the disease. Further research might involve analysis of different physiological signals and implementation of the developed framework into a real-life environment. Therefore, with this work we have come to complete all research objectives to address the problem with assigned enhancement.

All the Declarations and Statements

Author Contributions Statement

Sudhir Dhekane: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Visualization, Writing Original Draft Preparation.

Dr. Anand Khandare: Supervision, Validation, Writing, Review & Editing, Project Administration. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

This research received no specific grant from any funding agency in the public, commercial, or not-for profit sectors.

Data Availability Statement

No datasets were generated during the current study. Dataset analysed in this study is obtained from
EEG Psychiatric Disorders Dataset

Ethical Declarations

All procedures adhered to ethical guidelines for research.

Acknowledgments

The authors would like to express sincere gratitude to the Department of Computer Engineering, Thakur College of Engineering and Technology, for the invaluable support and resources provided throughout this research. The facilities, academic environment, and encouragement from faculty members have significantly contributed to the completion of this work. This study would not have been possible without the institution's commitment to advancing research and innovation in the field of brain computer interface

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used generative AI and AI-assisted technologies solely for language refinement, grammar correction, and improvement of readability. The authors carefully reviewed and edited the generated output as needed and take full responsibility for the content of this publication.

References

[1] "Monintoring Student Depression by Mobile Devices with Autoencoder Networks," Feb. 2023, doi: 10.1109/nnice58320.2023.10105715.

[2] P. Dai et al., "Classification of recurrent major depressive disorder using a residual denoising autoencoder framework: Insights from large-scale multisite fMRI data.," Computer Methods and Programs in Biomedicine, vol. 247, p. 108114, Mar. 2024, doi: 10.1016/j.cmpb.2024.108114.

[3] S. V. Reddy, S. Goyal, T. K. Reddy, R. Vinjamuri, and J. Andreu-Pérez, "Riemannian deep feature fusion with autoencoder for MEG depression classification in smart healthcare applications," Elsevier BV, 2024, pp. 197–212. doi: 10.1016/b978-0-44-313233-9.00014-x.

[4] H.-G. Wang, Q.-H. Meng, L.-C. Jin, J.-B. Wang, and H.-R. Hou, "AMG: A Depression Detection Model with Autoencoder and Multi-Head Graph Convolutional Network," pp. 8551–8556, Jul. 2023, doi: 10.23919/ccc58697.2023.10240138.

[5] F. Noman et al., "Graph Autoencoders for Embedding Learning in Brain Networks and Major Depressive Disorder Identification," arXiv: Neurons and Cognition, Jul. 2021, [Online]. Available: http://arxiv.org/pdf/2107.12838.pdf

[6] Z. Zhao et al., "Automatic Assessment of Depression From Speech via a Hierarchical Attention Transfer Network and Attention

Autoencoders," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 2, pp. 423–434, Feb. 2020, doi: 10.1109/JSTSP.2019.2955012.

[7] H. Dibeklioglu, Z. Hammal, and J. F. Cohn, "Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, pp. 525–536, Mar. 2018, doi: 10.1109/JBHI.2017.2676878

[8] B. Jones, W. D. Taylor, and C. G. Walsh, "Sequential autoencoders for feature engineering and pretraining in major depressive disorder risk prediction," JAMIA open, vol. 6, Oct. 2023, doi: 10.1093/jamiaopen/ooad086.

[9] "Speech-based Depression Detection Using Unsupervised Autoencoder," 2022 7th International Conference on Signal and Image Processing (ICSIP), Jul. 2022, doi: 10.1109/icsip55141.2022.9886372.

[10] Alruwais, N., Alamro, H., Eltahir, M. M. E., Salama, A. S., Assiri, M., & Ahmed, N. A. (n.d.). Modified arithmetic optimization algorithm with Deep Learning based data analytics for depression detection. AIMS Mathematics. https://doi.org/10.3934/math.20231549

[11] Su, M.-H., Wu, C.-H., Huang, K.-Y., & Yang, T.-H. (2020). Cell-Coupled Long Short-Term Memory With L -Skip Fusion Mechanism for Mood Disorder Detection Through Elicited Audiovisual Features. IEEE Transactions on Neural Networks, 31(1), 124–135. https://doi.org/10.1109/TNNLS.2019.2899884

[12] Wang, S. (2024). An Applied Study of Multi-modal Emotion Recognition to Assist in The Diagnosis of Depression. Highlights in Science Engineering and Technology, 119, 454–460. https://doi.org/10.54097/y910db53

[13] Barkat, Y., Hamitouche, D., Parekh, D., Guo, I., & Benrimoh, D. (2025). Latent Space Data Fusion Outperforms Early Fusion in Multimodal Mental Health Digital Phenotyping Data. arXiv.Org, abs/2507.14175. https://doi.org/10.48550/arxiv.2507.14175

[14] Gerych, W., Agu, E., & Rundensteiner, E. A. (2019). Classifying Depression in Imbalanced Datasets Using an Autoencoder- Based Anomaly Detection Approach. IEEE International Conference Semantic Computing, 124–127. https://doi.org/10.1109/ICOSC.2019.8665535

[15] W. Gerych, E. Agu, and E. A. Rundensteiner, "Classifying Depression in Imbalanced Datasets Using an Autoencoder- Based Anomaly Detection Approach," IEEE International conference semantic computing, pp. 124–127, Jan. 2019, doi: 10.1109/ICOSC.2019.8665535.

[16] S. A. Fields, J. Schueler, K. M. Arthur, and B. Harris, "The Role of Impulsivity in Major Depression: A Systematic Review", doi: 10.1007/s40473-021-00231-y/Published.

[17] Z. Li, M. Ruan, J. Chen, and Y. Fang, "Major Depressive Disorder: Advances in Neuroscience Research and Translational Applications," Jun. 01, 2021, Springer. doi: 10.1007/s12264-021-00638-3.

[18] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9 Validity of a Brief Depression Severity Measure."

[19] P. Riera-Serra et al., "Clinical predictors of suicidal ideation, suicide attempts and suicide death in depressive disorder: a systematic review and meta-analysis," Eur Arch Psychiatry Clin Neurosci, Oct. 2023, doi: 10.1007/s00406-023-01716-5.

[20] P. W. Corrigan, B. G. Druss, and D. A. Perlick, "The impact of mental illness stigma on seeking and participating in mental health care," Psychological Science in the Public Interest, Supplement, vol. 15, no. 2, pp. 37–70, Jan. 2014, doi: 10.1177/1529100614531398.

[21] S. Olbrich and M. Arns, "EEG biomarkers in major depressive disorder: Discriminative power and prediction of treatment response," International Review of Psychiatry, vol. 25, no. 5, pp. 604–618, 2013, doi: 10.3109/09540261.2013.816269.

[22] B. Ay et al., "Automated Depression Detection Using Deep Representation and Sequence Learning with EEG Signals," J Med Syst, vol. 43, no. 7, Jul. 2019, doi: 10.1007/s10916-019-1345-y.

[23] G. M. Bairy, U. C. Niranjan, and S. D. Puthankattil, "AUTOMATED CLASSIFICATION OF DEPRESSION EEG SIGNALS USING WAVELET ENTROPIES AND ENERGIES," J Mech Med Biol, vol. 16, no. 3, May 2016, doi: 10.1142/S0219519416500354.

[24] Y. Mohammadi, M. Hajian and M. H. Moradi, "Discrimination of Depression Levels Using Machine Learning Methods on EEG Signals," *2019 27th Iranian Conference on Electrical Engineering (ICEE)*, Yazd, Iran, 2019, pp. 1765-1769, doi: 10.1109/IranianCEE.2019.8786540.

[25] S. Khan et al., "A machine learning based depression screening framework using temporal domain features of the electroencephalography signals," PLoS One, vol. 19, no. 3 MARCH, Mar. 2024, doi: 10.1371/journal.pone.0299127.

[26] P. Sandheep, S. Vineeth, M. Poulose and D. P. Subha, "Performance analysis of deep learning CNN in classification of depression EEG signals," *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, Kochi, India, 2019, pp. 1339-1344, doi: 10.1109/TENCON.2019.8929254.

[27] S. Soni, A. Seal, A. Yazidi, and O. Krejcar, "Graphical representation learning-based approach for automatic classification of electroencephalogram signals in depression," Comput Biol Med, vol. 145, p. 105420, 2022, doi: https://doi.org/10.1016/j.compbiomed.2022.105420.

[28] A. Sam, R. Boostani, S. Hashempour, M.

Taghavi, and S. Sanei, "Depression Identification Using EEG Signals via a Hybrid of LSTM and Spiking Neural Networks," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 4725–4737, 2023, doi: 10.1109/TNSRE.2023.3336467.

[29] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, and O. Krejcar, "DeprNet: A Deep Convolution Neural Network Framework for Detecting Depression Using EEG," IEEE Trans Instrum Meas, vol. 70, 2021, doi: 10.1109/TIM.2021.3053999.

[30] Y. L. Neo, "A Review of State-of-the-Art Machine Algorithms on Classifying Depressive Disorder EEG Signals," in 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2024, pp. 365–371. doi: 10.1109/ICAICA63239.2024.10822999.

[31] D. Jan, M. de Vega, J. López-Pigüi, and I. Padrón, "Applying Deep Learning on a Few EEG Electrodes during Resting State Reveals Depressive States: A Data Driven Study," Brain Sci, vol. 12, no. 11, Nov. 2022, doi: 10.3390/brainsci12111506.

[32] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," Comput Methods Programs Biomed, vol. 161, pp. 103–113, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.012.

[33] D. Dhabliya, K. S. Bhuvaneshwari, P. Kapoor, S. Sowmiya, A. Garg, and L. Yashoda, "Hybrid ANFIS Model For Depression Recognition Consuming Three-Channel EEG Data," in 2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), 2024, pp. 1–5. doi: 10.1109/ICRASET63057.2024.10895792.

[34] U. R. Acharya et al., "A novel depression diagnosis index using nonlinear features in EEG signals," Eur Neurol, vol. 74, no. 1–2, pp. 79–83, Dec. 2015, doi: 10.1159/000438457.

[35] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, "A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis," J Neurosci Methods, vol. 358, Jul. 2021, doi: 10.1016/j.jneumeth.2021.109209.

[36] S. Dhekane and A. Khandare, "Applied Machine Learning in EEG data Classification to Classify Major Depressive Disorder by Critical Channels," Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 7, no. 3, pp. 581–596, Jul. 2025, doi: 10.35882/jeeemi.v7i3.719.

[37] J. Shen, J. Chen, Y. Ma, Z. Cao, Y. Zhang, and B. Hu, "Explainable Depression Recognition from EEG Signals via Graph Convolutional Network," in 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2023, pp. 1406–1412. doi: 10.1109/BIBM58861.2023.10386011.

[38] L. Kumar, K. Kaustubh, T. Rajkhowa, and S. A. Mattur, "EEG-ViLSTM: A Deep Learning Approach for Depression Detection Using EEG Signals," in 2025 National Conference on Communications (NCC), 2025, pp. 1–6. doi: 10.1109/NCC63735.2025.10983346.

[39]W. Yuan et al., "Discovery of Shared Latent Nonlinear Effective Connectivity for EEG-Based Depression Detection," IEEE Trans Neural Netw Learn Syst, vol. 36, no. 6, pp. 10663–10677, 2025, doi: 10.1109/TNNLS.2024.3514182.

[40] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbaf, "Major depressive disorder diagnosis based on effective connectivity in EEG signals: a convolutional neural network and long short-term memory approach," Cogn. Neurodyn., vol. 15, no. 2, pp. 239–252, A

[41] Zhang, Z., Yang, J., Xiong, P., Hao, H., Zhang, J., Li, L., Wang, C., & Liu, X. (2025). A cross subject MDD detection approach based on multiscale nonlinear analysis in resting state EEG. Neuroscience, 582, 1–10.

[42] S. M. Park et al., "Identification of Major Psychiatric Disorders From Resting-State Electroencephalography Using a Machine Learning Approach," Front Psychiatry, vol. 12, Aug. 2021, doi: 10.3389/fpsyt.2021.707581.

[43] Tigga, N. P., & Garg, S. (2022). Efficacy of novel attention-based gated recurrent units transformer for depression detection using electroencephalogram signals. Health information science and systems, 11(1), 1. https://doi.org/10.1007/s13755-022-00205-8

[44] Bairagi, V. EEG signal analysis for early diagnosis of Alzheimer disease using spectral and wavelet based features. *Int. j. inf. tecnol.* **10**, 403–412 (2018). https://doi.org/10.1007/s41870-018-0165-5

[45] S. Dhekane and A. Khandare, "Optimized EEG-Based Depression Detection and Severity Staging Using GAN-Augmented Neuro-Fuzzy and Deep Learning Models ", *j.electron.electromedical.eng.med.inform*, vol. 7, no. 4, pp. 1112-1129, Sep. 2025.